

EXHIBIT 1

Declaration of Meredith “Merry” McCarron in Support of Motion to Amend Case Schedule

**UNSEALED DOCUMENT
(ECF NO. 224-4)**

Lesley E. Weaver (SBN 191305)
Anne K. Davis (SBN 267909)
Joshua D. Samra (SBN 313050)
BLEICHMAR FONTI & AULD LLP
1330 Broadway, Suite 630
Oakland, CA 94612
Tel. (415) 445-4003
lweaver@bfalaw.com
adavis@bfalaw.com
jsamra@bfalaw.com

Gregory S. Mullens (admitted *pro hac vice*)
BLEICHMAR FONTI & AULD LLP
75 Virginia Road, 2nd Floor
White Plains, NY 10603
Tel. (415) 445-4006
gmullens@bfalaw.com

Plaintiffs' Interim Co-Lead Counsel

**UNITED STATES DISTRICT COURT
NORTHERN DISTRICT OF CALIFORNIA
SAN JOSE DIVISION**

In re Google Generative AI Copyright Litigation

Joseph R. Saveri (SBN 130064)
Cadio Zirpoli (SBN 179108)
Christopher K.L. Young (SBN 318371)
Elissa A. Buchanan (SBN 249996)
Evan A. Creutz (SBN 349728)
Aaron Cera (SBN 351163)
JOSEPH SAVERI LAW FIRM, LLP
601 California Street, Suite 1505
San Francisco, CA 94108
Telephone: (415) 500-6800
Facsimile: (415) 395-9940
Email: jsaveri@saverilawfirm.com
czirpoli@saverilawfirm.com
cyoung@saverilawfirm.com
eabuchanan@saverilawfirm.com
ecreutz@saverilawfirm.com
acera@saverilawfirm.com

Master File Case No. 5:23-cv-03440-EKL-SVK
Consolidated Case No. 5:24-cv-02531-EKL-SVK

**DECLARATION OF MEREDITH "MERRY"
McCARRON IN SUPPORT OF PLAINTIFFS'
MOTION TO AMEND CASE SCHEDULE**

1 I, Meredith “Merry” McCarron, hereby declare under penalty of perjury as follows:

2 1. I am a Co-Founder of MK Analytics, the company I co-founded in 2021 that leads the data
3 science and analytical efforts on high-stakes projects, especially in the context of litigation. I received my
4 Master’s Degree in Applied Informatics from New York University in 2014. I have held a number of highly
5 technical data science positions over the last 10 years, particularly roles that have allowed me to specialize
6 in complex litigation cases that involve Big Data challenges. I previously submitted a declaration in support
7 of Plaintiffs’ Motion to Amend Case Schedule, dated July 11, 2025 (ECF No. 174-1), which further
8 discusses my expert qualifications.

9 2. I make this declaration pursuant to 28 U.S.C. § 1746 in support of Plaintiffs’ Motion to
10 Amend Case Schedule.

11 3. I have been retained by Plaintiffs in this action to analyze the training data produced by
12 Defendants in this action in support of Plaintiffs’ class certification motion and intend to offer my opinions
13 as an expert in support of that motion.

14 4. In the ordinary course of my business, I am responsible for creating and maintaining the
15 physical and cloud computing resources for MK Analytics. I am familiar with and proficient in multiple
16 cloud computing platforms, including Google Cloud Platform (GCP) and Amazon Web Services (AWS). I
17 use one or both of these platforms on a daily basis, as does the MK Analytics staff, at my direction. My
18 extensive professional experience in this domain, especially my regular engagement with cloud computing
19 platforms, has positioned me well to recognize that Google has failed to provision Plaintiffs with sufficient
20 computing resources in a timely fashion.

21 5. I have extensive experience analyzing large datasets, including those produced by large tech
22 companies. The systematic and recurring stoppages, lockouts and data errors that have occurred in this case
23 are not routine. On a scale from 1 to 10, they are a 10. The persistent problems have consistently impeded
24 the workflow of my team and prevented us from doing the work we have been asked to do. We require, at
25 a minimum, six weeks of uninterrupted access to reliable data to complete our work plan.

26 **A. Unresolved Problems with the Training Data Review Environment**

27 6. On July 11, 2024, I submitted a Declaration in support of Plaintiffs’ Motion to Amend Case
28 Schedule (ECF No. 174-1) (“July 11 Decl.”), detailing the substantial challenges and delays that my team

1 and I experienced in attempting to access and analyze the training data that Google produced in this action.
2 As set forth in my July 11 Declaration, these challenges broadly encompassed: (i) significant delays in
3 Chromebook setup and configuration (July 11 Decl. ¶¶ 7-10), (ii) severe system performance problems that
4 rendered basic operations unusually slow (*id.* ¶¶ 11-14), (iii) technical limitations and missing critical
5 infrastructure that blocked industry-standard data processing tools and database access (*id.* ¶¶ 5, 14-17),
6 and data format issues that made the training materials disorganized and difficult to parse (*id.* ¶ 18).

7 7. Despite the two calls with Google’s technical team on July 28, 2025 and Aug 5, 2025, and,
8 from what I understand, a multitude of communications shepherded to Google’s technical team by
9 Plaintiffs’ counsel to Google on behalf of MK Analytics, many of the issues raised on July 11, 2024 remain
10 unresolved, and further issues introduced by Google’s control and management of the training data and
11 review environment have caused (and continue to cause) critical delays, which have made it virtually
12 impossible for MK Analytics to analyze Google’s training data in an expeditious fashion to meet the
13 Court’s class certification deadline of September 22, 2025.

14 8. I understand that Judge van Keulen held a hearing with the Parties on September 13, 2025,
15 to address Plaintiffs’ Motion to Enforce the Court’s Discovery Order to Produce Training Data, and that,
16 among other things, the Parties were to arrange a technical call between Google technician(s) and Plaintiffs’
17 experts. *See* Paragraph 22, *infra*.

18 **B. Delays Resulting from Inaccessible Training Data**

19 9. On July 29, 2025, my team spoke with Google’s engineers for the first time and learned for
20 the first time that we had been attempting to use a system (Google’s BigQuery) that was not and would not
21 be made available by Google. Google told us we should use a different program (TensorFlow) instead.
22 This revelation came as a complete surprise, particularly given that my July 11 declaration had specifically
23 indicated that I did not have access to the BigQuery tables of training data set up by Google (July 11 Decl.
24 ¶¶ 17-18), clearly communicating MK Analytics’ expectations and understanding that the training data had
25 been produced in BigQuery format. Despite this clear indication of our expectations, Google did not take
26 action to inform MK Analytics that the training data had not been produced in BigQuery and must be
27 analyzed with TensorFlow until July 29—18 days after the submission of my declaration—and only when
28 I was able to interface directly with some of Google’s engineers. The TensorFlow program also would not

1 work properly on the laptops Google had given MK Analytics. On top of that, much of the data Google
2 had produced was stored in SSTable format, a format that could not open or read with TensorFlow in
3 Google's environment. As a result, Google had to spend days converting these training datasets into formats
4 that TensorFlow could read and analyze, which Google did not provide until August 9, 2025.

5 10. When Google provided the training data in a form that could be analyzed by the tools
6 Google provided (TensorFlow), my team was still unable to access the complete training data to conduct
7 our analysis. On August 13, MK Analytics identified 31 datasets that were either corrupted or inaccessible,
8 which my team notified Plaintiffs' counsel of the same day. As result, MK Analytics could not access these
9 31 datasets until August 21, at which point Google had communicated that the issues with the datasets had
10 been resolved, which as I described below in Paragraphs 12-14 herein, ultimately proved to be untrue.

11 **C. Severe Performance Problems and Outages with Training Data Environment**

12 11. Since July 1, 2025, my team and I have cumulatively lost 124 working hours as result of 19
13 different system outages and frozen environments events. Each time one of the MK Analytics team
14 members loses access to Google's environment due to system outages or frozen environments, we must
15 wait for a Google engineer to reset the environment before we regain access. Due to the environment
16 rebooting, code often breaks in the middle of the night, causing us to lose important data processing time
17 because the code must be restarted the next morning. Code that should have been run in approximately 23
18 hours has needed two and a half days due to the environment reboot issues introduced by Google's set-up
19 configuration. Plaintiffs' experienced experts rate the ongoing issues, on a scale from one to ten, at a ten in
20 terms of an impossible work environment.

21 **D. Delays in Production of Work Product from the Review Environment**

22 12. I have made three "takeout" requests from the review environment, comprising my work
23 product, that have been untimely produced by Google or subject to delayed processing due to technical
24 limits that were previously undisclosed by Google and, in my experience, are unreasonable.

25 13. My takeout request, dated August 28, 2025, was untimely produced to Plaintiffs' counsel
26 (2 days after the deadline for production). This total takeout request was 11 GB, and could have been
27 downloaded in a day and produced by Google but was not.
28

14. Another takeout request, September 2, 2025, was initially not produced because Google deemed it “excessive” and, as I understand it, required attorney-review by counsel for Google before it was produced. This takeout request was 600GB, approximately .002% of the training data in the review environment, and could have been downloaded and produced by Google in a day but it was not. In my experience in litigations working with takeouts from remote environments, it is unprecedented for opposing counsel to review the takeout, *i.e.*, the expert’s work product, and refuse to produce it. I understand that on September 11, 2025, Google withdrew its objections and produced this takeout on September 12, 2025.

15. My takeout request, September 11, 2025, was the subject of processing delay due to a takeout limit that was previously unknown to me and not disclosed by Google. Specifically, Google informed Plaintiffs’ counsel on September 11, 2025, after my takeout request was submitted, that the review environment would not permit any takeout download if more than ten (10) files were present in the takeout folder at one time. This was the first time Google informed of any file-limit restriction. To attempt to cure the problem, and at Google’s request, I removed other takeouts that I had placed in queue for production to bring the file number below 10. This did not cure the problem. I understand that as of September 12, 2025, Google represented that the takeout limit will be increasing, but the increase has not been implemented yet and there is no date certain for that increase.

E. Recommendations to Resolve Problems with the Training Data Review Environment

16. Based on the extensive delays and technical challenges documented above, MK Analytics recommends the following immediate actions to remedy the ongoing issues with the training data review environment and Plaintiffs’ access to it:

a) **Eliminate the 24-hour timeout restriction.** The current process limitation, which times out at irregular intervals, is incompatible with the scale of data analysis required. Processing terabytes of training data necessitates continuous operations that often run for days. The environment should support uninterrupted processing for at least 72 hours, with a regular re-set schedule, to enable meaningful analysis of large datasets with predictable restarts.

b) **Permit user reboots for system freezes.** Currently, system outages and frozen environments require waiting hours or sometimes days for Google’s engineers to reset

1 the environment. Permitting users to reboot the system when the environment freezes
2 would potentially alleviate such delays. In the alternative, a dedicated technical support
3 team should be available with reasonable response times.

4 c) **Allow self-service data takeouts.** Under the current protocol, MK Analytics must
5 request each data takeout through counsel, then Google has three business days to
6 produce the requested data. In my experience, self-service takeouts operate efficiently
7 in similar review environments, without a three-day waiting period.

8 17. The aforesaid recommendations represent the minimum necessary steps to create a
9 functional analysis environment that would provide reasonable confidence that MK Analytics can complete
10 its analysis of the training data without further delays stemming from the systemic issues described herein.
11 In my experience, a remote environment administered by a third-party provider has not been saddled with
12 nearly as many chronic access issues as the one currently at issue in this case.

13 18. I am currently working as an expert on a litigation in which a shared third-party cloud-based
14 coding environment is in place, and the data at issue is of comparable complexity and none of the system
15 freezing and reboot issues have been present, nor is a 24-hour authentication timeout imposed
16 causing long-running code to break. Further, in that case plaintiffs are able to extract subsets of data on
17 their own without having to go through an approval process from the defense.

18 **F. Google's Recent Changes and Additions to the Training Data**

19 19. On Sunday, August 31, 2025, I learned that the day before, Google had disclosed for the
20 first time that when it converted the training data from one file format to another, it left out important
21 identifying information—specifically, web addresses (URLs) that showed where the data originally came
22 from. Google informed Plaintiffs' counsel that Google needed to redo the conversion process to include
23 this missing information and add these corrected files to our review environment. It is my understanding
24 that Google did not disclose at that time the scope of the datasets that were impacted and that would need
25 to be reconverted. As result, myself and MK Analytics were left with no information regarding scope of
26 what datasets were affected, what datasets could continue to be analyzed due to being unaffected, and
27 which datasets my team should cease analyzing.

28 20. On Tuesday, September 2, 2025, Google disclosed for the first time that 49 of the 60 datasets

Google produced were impacted by conversion issues. Google revealed that 31 datasets contained web address URLs that had been omitted, and additionally disclosed for the first time that the DocID field—which appears to be a critical identifier Google used to track the provenance and organization of works contained in the training data—was also missing from certain datasets. This DocID field may be of critical importance to my team’s analysis and the opinions I will render in support of class certification.

21. The size of the 49 datasets, which Google refers to as “files,” is enormous. It amounts to 600,000 records reflecting billions of items or works. While the Parties use different definitions for the term dataset and file, Google has not contested that the 49 datasets (or files) at issue amount to 600,000 records reflecting billions of items or works.

22. It was not until Wednesday, September 3, 2025, that I learned that Google had completed the update to the environment. At that time, my team at MK Analytics and I immediately began analyzing the changes and additions to the training data environment that Google made. This included, among other things, assessing for each dataset what information was previously unavailable, to what extent it may implicate our developed methodology, and to what extent the results of our analysis to date have been affected.

23. On September 12, 2025, I, along with a member of my expert team from MK Analytics, spoke with technical personnel for Google regarding the recent changes to data that required my re-analysis over the last twelve days (at the expense of conducting other expert analysis). Based upon what was told to me regarding the scope of the change, I am now able to move forward with my expert analysis of the impacted data. The information provided by Google on today’s call was not available to me before today’s call, nor could I have ascertained the information independently.

G. Current Impact Assessment of Google’s Changes to the Training Data Environment

24. Google’s changes to the training data environment have required MK Analytics to devote further time and analysis to determine the impact on its work to-date and the completion of its analyses from the additional training data element Google has recently produced. MK Analytics has devoted significant time solely to assessing and addressing the impact of Google’s data conversion errors, rather than conducting our substantive analysis of the training data itself.

25. Assuming the issues described above are resolved (and do not resurface), accounting for the

1 time lost (as set forth above), and based upon my experience, an aggressive estimate is that it will take at
2 least six weeks from the current class certification deadline of September 22, 2025, to: (1) complete my
3 reanalysis of data that was conducted before and during modifications of that data by Google; (2) complete
4 my analysis of data that was put on hold pending my reanalysis of modified data; (3) continue and complete
5 testing of takeout material; and (4) prepare and author expert reports.

6 *****

7 I declare under penalty of perjury under the laws of the United States of America that the foregoing
8 is true and correct. Executed this 12th day of September, 2025, at Philadelphia, Pennsylvania.

9 /s/ Merry McCarron
10 Meredith "Merry" McCarron
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28

ATTESTATION PURSUANT TO CIVIL LOCAL RULE 5-1(i)(3)

I, Lesley E. Weaver, attest that concurrence in the filing of this document has been obtained from the other signatory. I declare under penalty of perjury that the foregoing is true and correct.

Executed this 12th day of September, 2025, at Oakland, California.

/s/ Lesley E. Weaver

Lesley E. Weaver